엔진에 대한 품질보증데이터의 비모수적 분석 사례연구

백재욱*† · 조진남**

* 한국방송통신대학교 정보통계학과
** 동덕여자대학교 정보과학대학 데이터정보전공

Nonparametric Analysis of Warranty Data on Engine : Case Study

Jaiwook Baik*† · Jinnam Jo**

* Department of Information Statistics, Korea National Open University
** Department of Data & Information Science, Dongduk Women's University

Key Words : Warranty Data, Population Formation, Renewal Data, Hazard Rate

Abstract

Claim history data of rather long period were collected to assess reliability and warranty cost analyses. The data were appropriately organized to be used for further statistical analyses. For each critical component, nonparametric statistical method was applied to obtain reliability plot. Hazard plots of the components in a subsystem or system level were also obtained. Competing risk model was assumed to obtain the performance of the subsystem or system level.

1. Introduction

Quality assurance programs are becoming popular. Some of the sectors in the automobile industry have begun warranty systems in which every new car is warranted for an extended period of time. Especially some parts of the automobile such as engine and transmission are under warranty for more than 5 years, which means that people are reporting their problems with their cars for much longer period of time than before. Therefore manufacturers now have an access to rather reliable warranty data of long history. The problem is how the manufacturers can use the warranty data so that any valuable information can be extracted from them. A key concern would be the estimation and control of reliability and warranty cost elements associated with warranty [4, 6]. Research has been done on warranty data analysis [5, 7, 8, 15].

Several different types of analyses can be formulated ; one-dimensional or two-dimensional analysis, analysis assuming renewal or non-homogeneous Poisson process, and component level or subsystem (or system) level analysis. For instance, we can focus on subsystem (or system) level analysis assuming minimal repair (nonhomogeneous Poisson process) in one or two dimensions. Rai and Singh [13] find interesting results modeling mileage in addition to age. See also Rai and Singh [12] and Majeske [10].

Various models have been proposed for re-

† 교신저자 jbaik@knou.ac.kr

pairable systems. Asher and Feingold [2] reviewed some ofthese models, including homogeneous Poisson process (HPP), nonhomogeneous Poisson process (NHPP) and renewal process. HPP can be used if the distribution of time between failures is assumed to be an exponential distribution. HPP is a special type of renewal process in which the inter-arrival times have an exponential distribution with the same renewal rate throughout the observed period. This simple model may not be appropriate in mechanical components since there may be a change in the hazard rate over time.

NHPP is a good model for a repairable system because it can model systems that are deteriorating or improving over time. Especially, it is appropriate in mechanical components since the system is deteriorating after repair. Baik, Murthy and Jack [3] dealt with NHPP with two-dimensions. In this case study, however, competing risk model is used to model the behavior of the subsystem and system levels for one dimension.

We optimistically assume that each component after repair (or replacement) functions a new and we focus on one-dimensional (namely age) renewal process at the component level and use the result in the subsystem (or system) level analysis. We deal with completely observed process of claim history [9, 11].

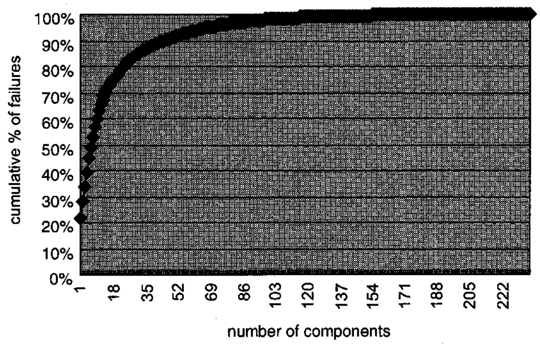
In this case study the warranty data on engine were collected in an Excel format and given to data analyst. We show how to identify critical components in engine from the raw warranty data, and how to form a conceptual population for each critical component and perform reliability analysis. In addition, since a subassembly (subsystem) is composed of several critical components and since each assembly (system) is again composed of several subassemblies we show how to perform the reliability analysis at the subassembly and assembly levels. These types of reliability analysis can help the management with the decision on how to improve components, subassemblies and assembly, and reduce the warranty cost.

This case study starts with raw claim data collected from the A/S centers and handed over to the data analyst. The data are incomplete in the sense that some information is missing and some are even incorrect. In order for the analyses to be valid the data itself have to be reliable. Therefore, in Section 2 the raw claim data received are screened out so that only relevant data for the study could remain. Then critical components are identified in terms of the number of claims. Last of all, critical components are grouped into several categories (subassemblies). In Section 3 a valid conceptual population is formed for each critical component assuming that the car is renewed every time repair or replacement takes place. Analyses at the component, subassembly and assembly levels will be performed in later sections with the conceptual population. In forming a population for each component the cars that have been sold but not been claimed are used to generate censored lives. Claimed cars have also been used to generate data since they produce failure times as well as censored times. In the latter case censored times are calculated as the repair or replacement time (last repair or replacement time if there were more than one claim on the same component in a car) to the data collection time. In Section 4 reliability analysis is performed at the component level. Simple parametric models such as Weibull and Lognormal distributions are tried to the appropriate data at the component level. However, they turned out to be irrelevant to the current warranty data of long history. Therefore, nonparametric methods are used to obtain reliability and hazard rate functions. In Section 5 analyses at the subassembly and assembly levels are performed based on the analyses at the component levels assuming independence among components failures and competing risk model. Specifically hazard plot for each subassembly can be obtained. Last of all, in Section 6 summary of what was found in this case study is given.

2. Data screening and critical component identification

In any company there are a lot of data : sales data, warranty data, and maintenance data to name a few. However, since a lot of them are observatory data in essence, not experimental data, it takes some time to sort out meaningful data needed for any analysis. Sometimes it is impossible to get any reliable data for the purpose. In this section, we will show how to screen raw data to find critical components in engine. First of all, we explain about our raw claim data. But for confidential reason, we will only introduce relevant data for the study. The claim data consist of 1 to 5 years of claim history with several columns of interest to the data analyst : vehicle identification number (VIN), production date (PD), sales date (SD), repair (or replacement) date (RD), component of the cause (CC), failure mode (FM), cause of the failure (CF), repaired (or replaced) component (RC), dealer code (DC). Therefore, for each claim VIN, PD, SD, RD, CC, FM, CF, RC and DC were recorded. Note that CC isthe component of the suspected or possible cause of the problem when the car came in to the A/S center while RC is actually the repaired or replaced component to rectify the claim. RC could be different from CC for some claims.

The data that had been handed over to the data analyst contained some irrelevant data for the study. For instance, some data belonged to some other types of cars that were of no interest for the study. In addition to the obvious irrelevant data, there were some dubious parts to them. More specifically, RC's were missing for more than half of the claim data. They could have been CC's. Or it could have been that there was just minor adjustment to CC's, or neither repair nor replacement. Thus, when we identify critical components in engine the claim data with missing RC's were ignored in order not to use incorrect information.



<Figure 1> Pareto analysis for component failures

The first descriptive statistics relates to frequency of failures for different components and the Pareto analysis is shown in <Figure 1>. It turned out that less than 20% of the components were responsible for almost 90% of the claims and so these components are termed "critical" components.

<Table 1> Partial Pareto analysis of failure

order	Component name	freq %	failure mode	freq %	cause of failure	freq %
1	CO_1	21%	FM_{11}	94%	CF_{111}	88%
					Others	12%
			Others	6%		
2	CO_2	6%	FM_{21}	57%	CF_{211}	34%
					CF_{212}	32%
					CF_{213}	15%
					Others	19%
			FM_{21}	34%	CF_{221}	56%
					CF_{222}	17%
					CF_{223}	16%
					Others	11%
			Others	9%		

Analysis for the failure modes and causes for the critical components were carried out and the results for the first two critical components are shown in <Table 1>. Each of the critical components can be categorized into several disjoint subassemblies with electronic and sealing subassemblies accounting for bulk of the claims. Since almost all of the critical components are under warranty except only a few we will assume that each of them contributes to the analysis at the subassembly or assembly level in Section 5.

3. Population formation

In this section, we are concerned about formation of a conceptual population for each critical component because we want to know how reliability changes as time goes by. In the previous section, claim data with missing RC's have been deleted when we identify critical components. However, it will overestimate (underestimate) the reliability (warranty cost) if we completely ignore them when we perform reliability and analysis. Therefore, we replace the missing RC's by CC's. However, in this case some of the rows in the database become identical. Then they have been treated as same claims.

Now we want to form an appropriate population for each component. Following Subsections 3.1 and 3.2 focus on how to get the number of censored cars for our warranty data, thereby censored lives to be added to the population. Subsection 3.3 focuses on how to extract lives (failure and censored)from the claim data. In Subsection 3.4 both data in the previous subsections are combined to form a population for each critical component.

3.1 Adjusted monthly number of cars sold based on claim data received

The data analyst received the actual monthly number of cars sold for a particular type of model from the manufacturer. However, we need to adjust the figures because the data analyst received only a fraction of the whole claim data from the manufacturer due to their reluctance to give away all of their claim data. In addition, since the data we received from the company were arranged in terms of production date with no record of sales date, the average time lag between production and sales was taken into account to get an adjusted data acquisition rate for each year. In some years the adjusted data acquisition rates were as low as 25%. We used the adjusted data acquisition rate to get the monthly-adjusted number of cars sold.

3.2 Monthly number of censored cars and censored lives

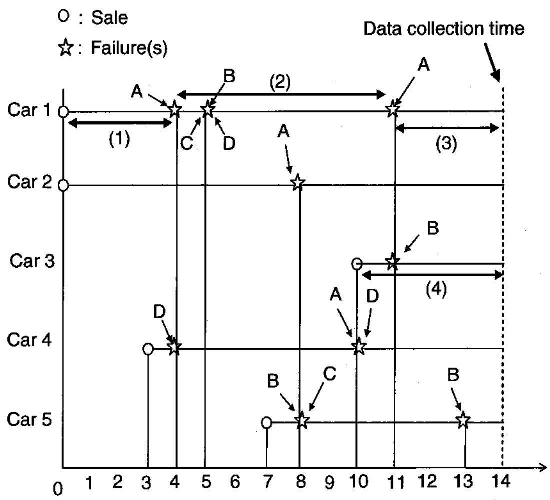
Censored lives from censored cars are easy to get in principle since they can be considered to be the sales date to the data collection date. Therefore, in our data first it is necessary to arrange the data in terms of the sales month. However, there were some data with sales dates of 01/01/1900 due to recording error. And some cars, if not many, even had sales dates of 01/01/2000 while the first repair or replacement was performed in 1999. Since there were not many data of those types we could have ignored them. But since we wanted to use all the information we had we used the mileages of those cars with correct sales dates to estimate the censored lives and included them in the database. The regression line drawn from the claim data with correct sales dates turns out to be 'Mileage = 49.3(miles) * Days'. However, note that the regression line was obtained only from claim data excluding censored lives coming from the censored cars.

Now we can get the monthly number of claimed cars from the above augmented database. And we can obtain monthly number of censored cars by subtracting the number of claimed cars from the adjusted monthly number of cars sold in 3.1. However, there were two suspicious parts in the monthly number of censored cars : data in the early 1996 and a couple of data in 1998. In the latter case, more cars than sold turned out to have been claimed. Since this does not make any sense we set the number of censored cars during those periods to be 0. In the former case, there seemed to be too many censored cars, especially in the early period of 1996. However, it should be the case since not many cars would have been sold within a few months since production (Note that the data analyst originally received warranty claim data ordered in terms of production month). So we need to know how long it took for a car to be sold since production. Then the new adjusted number of cars sold in early 1997 can be obtained. Finally for each censored car we can get the censored life as sales date to the data collection date and add it to the population of each critical component in the engine.

3.3 Failure and censored lives from the claimed cars for each component

Sample failures and censored lives that can be identified from the claimed carsat the component level are as shown in <Figure 2>. Specifically following 4 types of data can be extracted from the claimed cars :

a) For each claim data we get sales date to repair or replacement date as a failure data {See (1) of <Figure 2> for component A}.
b) If a component is replaced more than once in the same car we get previous repair or replacement date to recent repair or replacement date as a failure data {See (2) of <Figure 2> for component A}.
c) For any claimed car we get last repair or replacement date to data collection date as a censoring data {See (3) of <Figure 2> for component A}.
d) In some cars some components have never been repaired or replaced. In this case the cars generate only censored lives for those components. The censored lives in this case are taken as sales date to the data collection date {See (4) of <Figure 2> for component A}.



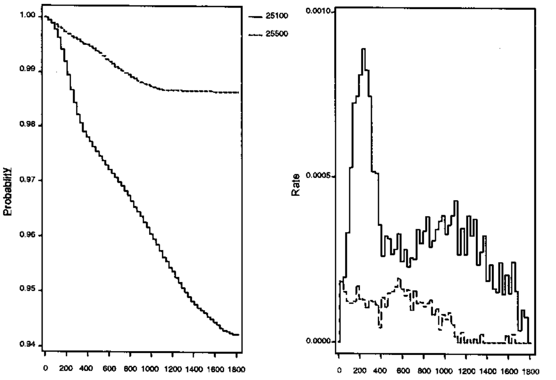
<Figure 2> Claim history in one dimension for 5 cars (Components A, B, C and D)

3.4 Combination of data

Now data from censored cars in 3.2 and from claimed cars in 3.3 are combined to make a conceptual population for each component and analyzed in the following sections; reliability analysis at the component level in Section 4 and at the subassembly and assembly levels in Section 5.

4. Analysis at the component level

Reliability analysis at the component level can be performed parametrically and nonparametrically for the data obtained in Section 3. For component 251**, there were a total of 35,766 data (=3,101 failure data + 32,665 censored data) from the claimed cars. In addition, 48,216 censored lives coming from censored cars were added to the population for the appropriate analysis of the component. Note that only 3.7% {=3,101/(48,216 + 35,766)} of the data were failures. This would be typical of the warranty data at the component level if we assume renewal or replacement at each failure of the car.



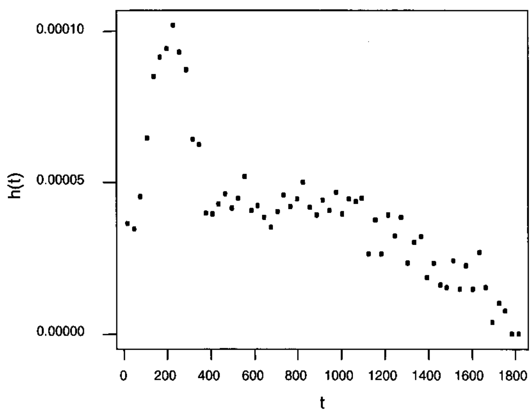
<Figure 3> Nonparametric analysis of components 251** and 255** in a subsystem (Reliability and hazard rate)

Weibull and Lognormal distributions are generally applied in the parametric method. However, in our warranty data it turned out that those distributions were not applicable since Anderson-Darling statistic is too large. Therefore, nonparametric methods are applied to the data. Minitab was used to draw the plots of reliability and hazard rates over time. For instance, the reliability and hazard plot of components 251** and 255** in a subsystem are as shown in <Figure 3>. Therefore, it can be said that components 251** and 255** are still working with probabilities of 95% and 99% even after 2.7 years (approximately 1000 days). Also it can be said that the probabilities that components 251** and 255** still working will fail at the age of 2.7 years are 0.003% and 0.001%, respectively.

Warranty cost analysis can be performed at component level since it can be repaired or replaced more than once over time in a car. As in the reliability analysis we use the nonparametric method for the warranty cost analysis. Especially, Nelson [11] can be used to draw a pictureof the mean cumulative number of failures for each component as time goes by (see also [14]). Then we can get the warranty cost for the component by multiplying the mean cumulative number of failures by the repair or replacement cost.

5. Analysis at the subsystem and system levels

The analysis of the subsystem or system depends on the nature of the data available as explained in Ansell and Phillips [1]. The data in this case study are available at component level. Hence we can combine all the information at component levels to make inference on the behavior of the subsystem. In this case study we assume that all of the components need to be working in order for the higher level of the system to function properly, therefore a series system.



<Figure 4> Nonparametric hazard plot for a subsystem with components 251** and 255**

<Figure 4> shows a hazard plot for a cooling subsystem. Note that since there are two critical components in the subsystem, the hazard rate for the subsystem is $h_{sub}(t) = h_1(t) + h_2(t)$. Therefore, it can be said that the probability that the cooling subsystem still working at 2.7 years will fail at that moment is 0.004%.

Finally, by combining all the information at the subsystem levels in engine we can identify the behavior of a series system with instantaneous repair : the performance of the system (engine) might be expressed as $h_s = \sum_{j=1}^{m'} h_j$ where m' is the total number of subsystems in the system. The hazard plot for the system is not shown in this paper due to confidential reason.

So far we have been concentrating on reliability analysis at subsystem and system levels. We can also draw the graph of warranty cost at subsystem and system levels once we know the repair or replacement cost of each component.

6. Summary

In this case study we were given claim history data and asked to do reliability and warranty cost analyses. First thing we did was to sort out unnecessary data that were not related to this study; data that did not belong to a particular type of vehicle and engine of interest were discarded. Then with only relevant data for the study we identified critical components in engine. Also for each critical component we found critical failure modes and for each failure mode we found critical causes of the failure. In this case study we found that component CO_1 was the most critical one. We also found that the most frequent failure mode for the component was FM_{11} while the most probable cause of the failure for the failure mode was CF_{111}.

For each critical component in engine we formed a conceptual population. The population consisted of censored data as well as failure data. Then we could perform reliability analysis at the component level. The analysis was done nonparametrically since the distributional assumptions did not hold for the data. We could draw graphs of reliability and hazard rate for each critical component.

The behavior of subsystem (or system) was assessed through its hazard function $h_{sub} = \sum_{i=1}^{m} h_i$ where h_i is the hazard function for component i and where m is the number of components in the subsystem (or system).

References

[1] Ansell J. I. and Phillips, M. J.(1994), *Practical Methods for Reliability Data Analysis*, Oxford University Press.

[2] Asher, H. and Feingold, H.(1984), *Repairable Systems Reliability, Inference, Misconceptions and their Causes*, New York, Marcel Dekker.

[3] Baik, J., D. Murthy, N. P., and Jack, N.(2004), "Two Dimensional Failure Modelling with Minimal Repair", *Naval Research Logistics*, Vol. 51, pp. 345-362.

[4] Blischke, W. R. and Murthy, D. N. P.(2000), *Reliability, Modeling, Prediction and Optimization*, John Wiley and Sons.

[5] Kalbfleisch, J. D. and Lawless, J. F.(1998), "Estimation of Reliability in Field-Performance Studies (with Discussion)", *Technometrics*, Vol. 30, pp. 365-388.

[6] Kalbfleisch, J. D., Lawless, J. F., and Robinson, J. A.(1991), "Methods for the Analysis and Prediction of Warranty Claims", *Technometrics*, Vol. 33, pp. 273-285.

[7] Lawless, J. F.(1998), "Statistical Analysis of Product Warranty Data", *International Statistical Review*, Vol. 66, pp. 40-60.

[8] Lawless, J.F., Hu, J., and Cao, J.(1995), "Methods for Estimation of Failure Distribution and Rates from Automobile Warranty Data", *Life Data Analysis*, Vol. 1, pp. 227-240.

[9] Lawless, J. F. and Nadeau, J. D.(1995), "Some Simple Robust Methods for the Analysis of Recurrent Events", *Technometrics*, Vol. 37, pp. 158-168.

[10] Majeske, K. D.(2003), "A Mixture Model for Automobile Warranty Data", *Reliability Engineering System and Safety*, Vol. 81, pp. 71-77.

[11] Nelson, W.(1998), "An Application of Grap-

hical Analysis of Repair Data", *Quality and Reliability Engineering International*, Vol. 14, pp. 49-52.

[12] Rai B. and Signh, N.(2003), "Hazard Rate Estimation from Incomplete and Unclean Warranty Data", *Reliability Engineering System and Safety*, Vol. 81, pp. 79-92.

[13] Rai, B. and Signh, N.(2004), "Modeling and Analysis of Automobile Warranty Data in Presence of Bias Due to Customer-Rush Near Warranty Expiration Limit", *Reliability Engineering System and Safety*, Vol. 86, pp. 83-94.

[14] SAS Institute Documentation for the RELIABILITY PROCEDURE (available from J. Johnston, SAS Institute, SAS Institute Drive, Cary, NC 27513, USA), 1997.

[15] Suzuki, K., Karim, M. R., and Wang, L. (2001), Statistical Analysis of Reliability Warranty Data (Chapter 21 in Balakrishman, N and Rao, C.R. (eds), Handbook of Statistics, Elsevier Science, Amsterdam).